Week 22 : Hypothesis Testing

pending tasks	
type	

There are 8 generic steps in hypothesis testing :

State H_1	a bold claim		
State H_0	the opposite of H_1		
Collect a sample (n)	the larger the better		
Compute mean	say, $\bar{x}$		
Compute test statistic	$z = \dfrac{\bar{x} - \mu_0}{\sigma/\sqrt{n}}$		
Decide α	$\alpha = 0.01, 0.05, 0.1$		
Apply decision rule:	$\lvert z \rvert > z_{\frac{\alpha}{2}}$	$z > z_\alpha$	$z < -z_\alpha$
	$H_1 : \mu \neq \mu_0$	$H_1 : \mu > \mu_0$	$H_1 : \mu < \mu_0$

7 steps of hypothesis testing

1. State H_1 : it is the bold claim being made that has to be proven. Also known as the **alternate hypothesis,** it is usually the opposite of the null hypothesis.
2. State H_0 : also known as **null hypothesis,** it is the general belief that exists against which the alternate hypothesis is stated.
3. Collect a sample (n) : a sample with size n is collected, larger size is better.
4. Compute mean (or any other statistic)
5. Compute test statistic : To standardize the statistic. In case of mean a normal/t distribution is used depending on the availability of σ. For proportion a normal distribution is used and to prove independence chi square distribution is used.
6. Decide α : this decides the leniency towards selection of alternate hypothesis. Lesser the value, stricter the rule to reject null hypothesis H_0.
7. Apply decision rule: if the computed statistic lies in the region of α, the null hypothesis is rejected else it is accepted.
8. Compute p-value : it is the probability of occurrence of the alternate hypothesis. If it is lesser than α, the null hypothesis is rejected.

This gives rise to 4 types of problems that only differ in computation of the statistic :

	Mean	Proportion	Independence
known variance	use normal dist.	use normal dist.	use Chi sq. dist.
unknown variance	use t dist. (df = n-1)		

Fig.1 Types of problems in hypothesis testing.

Mean (known variance) - 6 case studies

- The mean of the sample follows certain normal distribution, it is possible to take sample with a different mean value,therefore, the evidence thus obtained is not enough the null hypothesis H_0.
- The result of an extreme sample is considered and H_0 is rejected only when its probability of occurrence is very low since, this implies that the population perhaps follows a different distribution(has different mean).
- A threshold probability α is selected to quantify the extreme value beyond which if the sample mean occurs H_0 is rejected.

Case study 1. A company selling chips claims that each packet has 200gm of chips. You are skeptical of the claim and believe that on average each packet does not contain 200gm of chips. Prove your claim if the variance is 10gm and tolerance limit is 5%.

H_1 : $\mu \neq$ 200 (bold claim)

H_0 : μ = 200 (status quo)

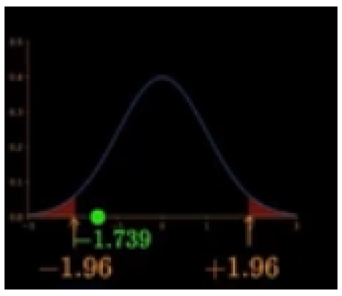
Sample : [193,212,174,200,195,195,194,198,181,203]

Mean $\bar{x}$ = 194.5

Test statistic $z = \frac{\bar{x}-\mu_0}{\sigma/\sqrt{n}} = \frac{194.5-200}{10\sqrt{10}} = -1.739$ (σ = 10 is known)

α = 0.05

Decision rule : reject H_0 if $|z| > z_{\alpha/2}$
but | -1.739| < 1.96



The null hypothesis cannot be rejected based on this sample as it is not extreme enough. The probability of occurrence of such a sample is given by p-value.

p value = $P(|z| > \bar{x}) = 0.082$

8.2% chance of occurrence, which is greater than tolerance limit of 5%.

Case study 2 : A company manufacturing ball bearings claims that the average radius of the ball bearings is 3mm. Your company purchases these ball bearings and has asked you to verify the claim if the variance is 0.5mm.

H_1 : $\mu \neq$ 3 (bold claim)

H_0 : μ = 3 (status quo)

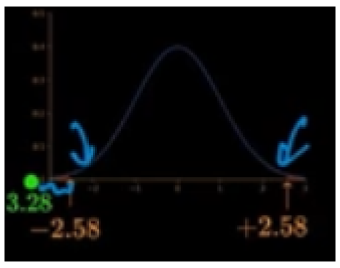
Sample : [2.99,2.99,2.70,2.92,2.88,2.92,2.82,2.83,3.06,2.85]

Mean $\bar{x}$ = 2.896

Test statistic $z = \frac{\bar{x}-\mu_0}{\sigma/\sqrt{n}} = \frac{2.896-3.0}{0.1\sqrt{10}} = -3.28$ (σ = 10 is known)

α = 0.01

Decision rule : reject H_0 if $|z| > z_{\alpha/2}$
but | -3.28| < 2.58.



The null hypothesis is rejected as the sample mean is extreme compared to that of population.

p value = $P(|z| > \bar{x}) = 0.001$

Both the extremes are not accepted in the case study 1 and 2, therefore a two-tailed test is done.

Case study 3 : You have developed a new dialog stem and done a user study. You claim that the average rating given by the users is greater than 4 on a scale of 1 to 5. How do you prove this to your critiques? (use deviation = 0.5)

H_1 : $\mu \geq 4$ (bold claim)

H_0 : $\mu \leq 4$ (status quo)

Sample : [4,3,5,4,5,3,5,5,4,2,4,5,5,4,4,5,4,5,4,5]

Mean $\bar{x}$ = 4.25

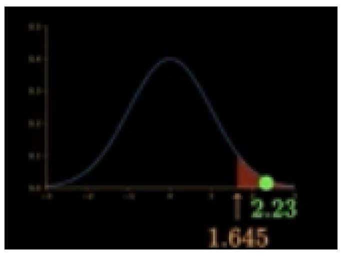
Test statistic $z = \frac{\bar{x}-\mu_0}{\sigma/\sqrt{n}} = \frac{4.25-4}{0.5\sqrt{20}} = 2.23$ (σ = 0.5 is known)

α = 0.05

Decision rule : reject H_0 if $z > z_{\alpha}$
but 2.23 > 1.645

The null hypothesis is rejected.

p value = $P(z > \bar{x}) = 0.012$



The p values indicates that there is only 1.2% chance to come across such a sample.

Case study 4 : You have developed an AI powered fuel management system for SUVs. You claim that with this system, on an average the SUVs' mileage is at least 15km/liter. Prove your claim if deviation =1.

$H_1 : \mu \geq 15$ (bold claim)

$H_0 : \mu \leq 15$ (status quo)

Sample=[14.08 ,14.13, 15.65, 13.78, 16.26, 14.97, 15.36, 15.81, 14.53, 16.79, 15.78, 16.98, 13.23, 15.43, 15.46, 13.88, 14.41, 14.31, 15.76, 15.38]

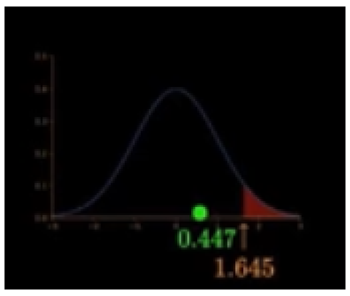
Mean $\bar{x}$ = 15.1

Test statistic $z = \frac{\bar{x}-\mu_0}{\sigma/\sqrt{n}} = \frac{15.1-15}{1\sqrt{20}} = 0.447$

α = 0.05

Decision rule : reject H_0 if $z > z_\alpha$
and 0.447 > 1.645

Even though the sample mean is greater than parameter, the sample considered is very likely thus, **the null hypothesis cannot be rejected.**



p value = $P(z > \bar{x}) = 0.327$

The case studies 3 and 4 discuss the case where the aim is to prove that statistic is greater than the parameter, thus one-tailed test is used.

Case study 5 : You have developed a new image classification system and claim that on average it takes less than 100ms to classify one image. How do you convince your reviewers about this claim? Standard deviation is 10ms.

$H_1 : \mu \leq 100$ (bold claim)

$H_0 : \mu \geq 100$ (status quo)

Sample n = 100

Mean $\bar{x}$ = 97.5

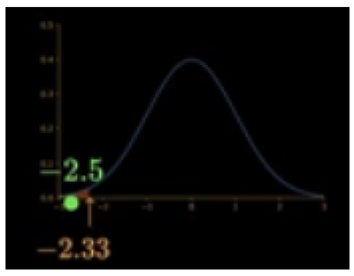
Test statistic $z = \frac{\bar{x}-\mu_0}{\sigma/\sqrt{n}} = \frac{97.5-100}{10\sqrt{100}} = -2.5$

α = 0.01

Decision rule : reject H_0 if $z < z_\alpha$
an -2.5 < -2.33

The null hypothesis is rejected.

p value = $P(z < \bar{x}) = 0.0006$



Case study 6 : You have developed a new machine translating system and claim that on average it takes less than 1 MB memory per sentence. How do you convince your reviewers about this claim? Based on the past data you know that the deviation is 0.1MB

$H_1 : \mu < 1$ (bold claim)

$H_0 : \mu \geq 1$ (status quo)

Sample n = 100

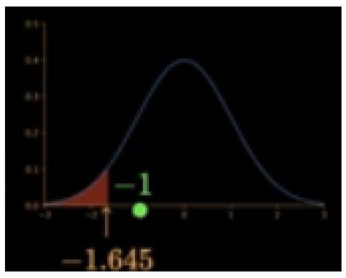
Mean $\bar{x}$ = 0.99

Test statistic $z = \frac{\bar{x}-\mu_0}{\sigma/\sqrt{n}} = \frac{0.99-1}{0.1\sqrt{100}} = -1$

α = 0.05

Decision rule : reject H_0 if $z < z_\alpha$
but -1 > -1.645

The null hypothesis cannot be is rejected as the example is not extreme enough.



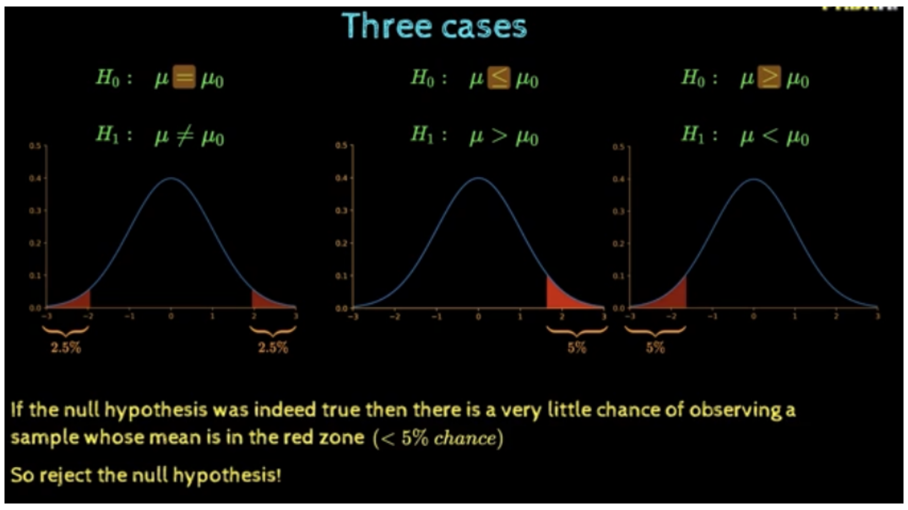
p value = $P(z < \bar{x}) = 0.158$

There is almost a 16% chance of observing such a sample.

Generalizing the Case Studies

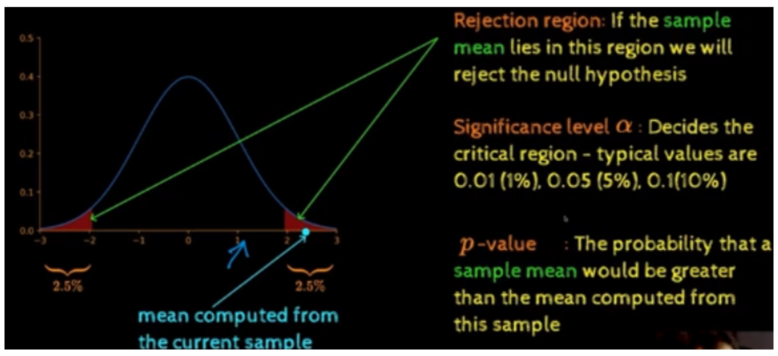
- The six case studies seen can be summarized into three high level cases. The null hypothesis and alternate hypothesis are opposite. When the null hypothesis has an inequality, the equality is considered to be the extreme case and the efforts are towards disproving this.

The three broad categories.



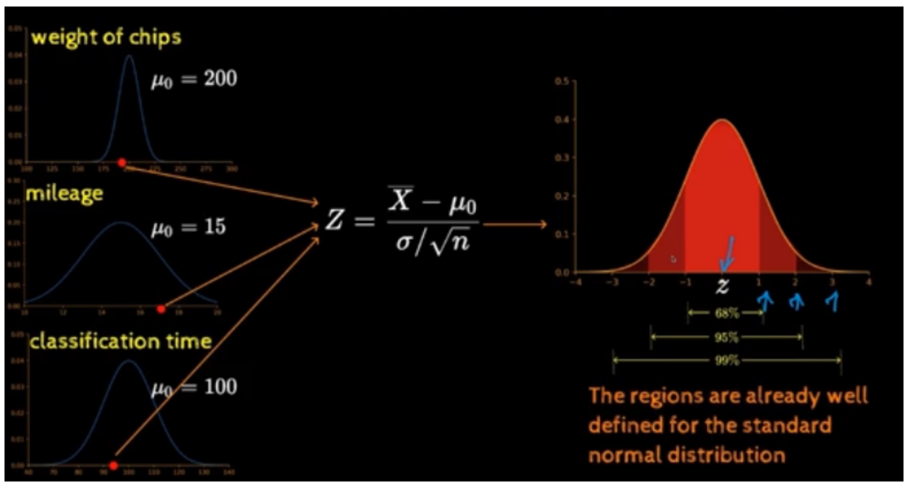
Area under the curve for each case.

- When the critical area (given by α) where sample mean must lie to reject the null hypothesis(**rejection region**) is split between the two tails it is called a 2-tailed test (used to prove $\neq$ cases) and, when the area is on either tails its called a 1-tailed test (for $\leq, \geq$ cases).
- **Significance level** α **:** decides the critical region, typical values are 0.01 (1%), 0.05(5%),0.1(10%). It is the tolerance level of null hypothesis acceptance (or leniency towards alternate hypothesis) and if the sample mean is beyond this value the null hypothesis is rejected.
- **p-value** : It is the probability that a sample mean would be greater than mean computed from the given sample.



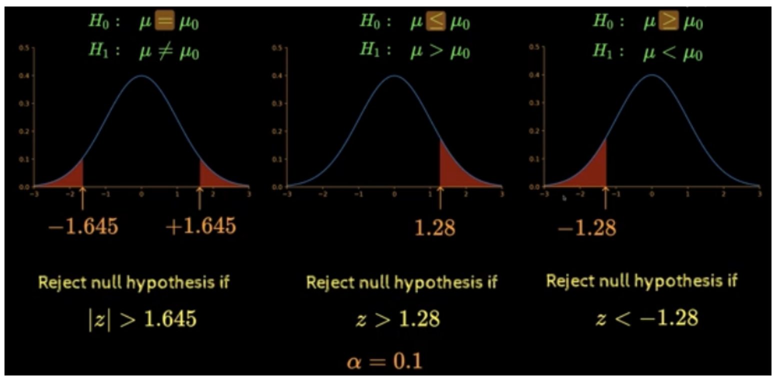
The sample mean can be replaced by any sample statistic.

- A normal distribution $N(\mu, \sigma/\sqrt{n})$ is mapped to $N(0,1)$ using $Z = \frac{\bar{X}-\mu_0}{\sigma/\sqrt{n}}$ for calculations, when the variance is known. This is called as **test statistic**.

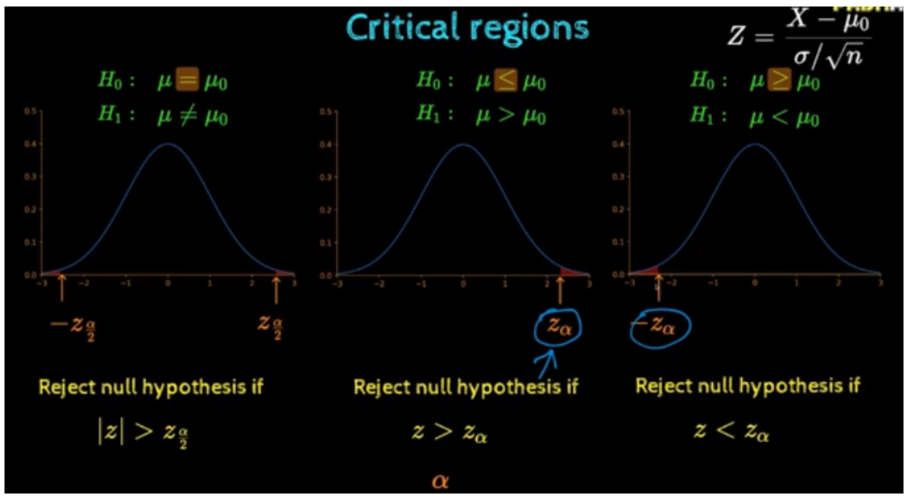


Mapping to N(0,1).

- $\alpha = 0.05$ corresponds to z value of $\pm$1.96 for two-tailed test and 1.645 (either +/-) for one-tailed test. If the statistic computed lies beyond these points, then the null hypothesis is rejected.



An example of a more strict rule for null hypothesis rejection.



Visualizing rejection region.

Effect of $n, \sigma \ and \ \alpha$

$$z = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} = \sqrt{n}\frac{\bar{x} - \mu_0}{\sigma}$$

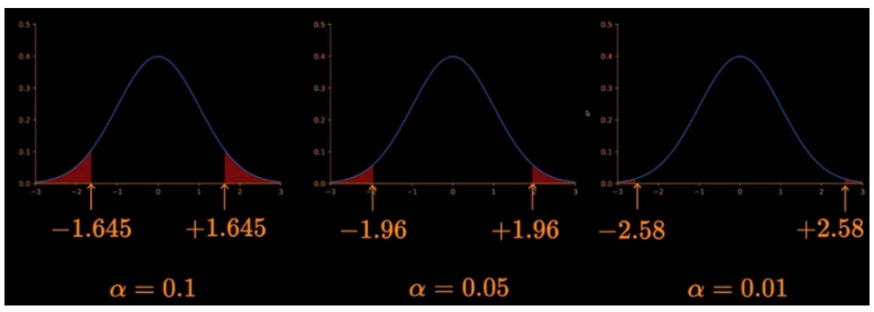
- If n is large the sample mean is expected to be closer to the population mean. **This implies that for a large sample even small differences (between the parameter and the statistic) would lead to rejection of null hypothesis**. The multiplication by $\sqrt{n}$ in the numerator ensures this. Increase in n implies bigger numerator and thus larger z values(which are lie near the tail). Smaller n would mean smaller multiplicative effect.

  In case study 1, a population size of 100 would result in z = -5.5 what would have been an extreme sample, thus, rejecting the null hypothesis.

- If σ is large, a lot of deviation is expected in sample mean from one sample to another. So small differences should not lead to rejection of null hypothesis. The division by σ ensures this.

  In case study 2, a deviation of 0.2 would have led to z = -1.64 and the null hypothesis would not have been rejected.

- As α is decreased, the rejection zone decreases. Lower the α, stricter is the requirement for rejecting null hypothesis.

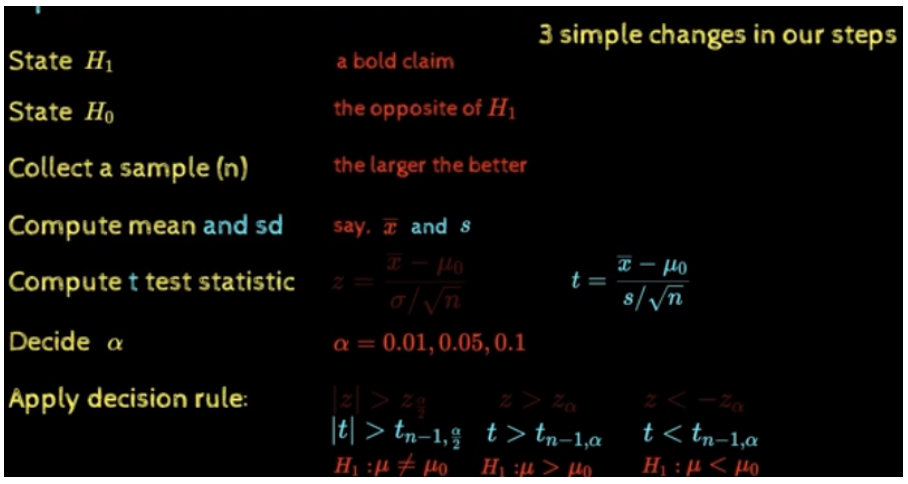


Reducing alpha results decreases he rejection zone.

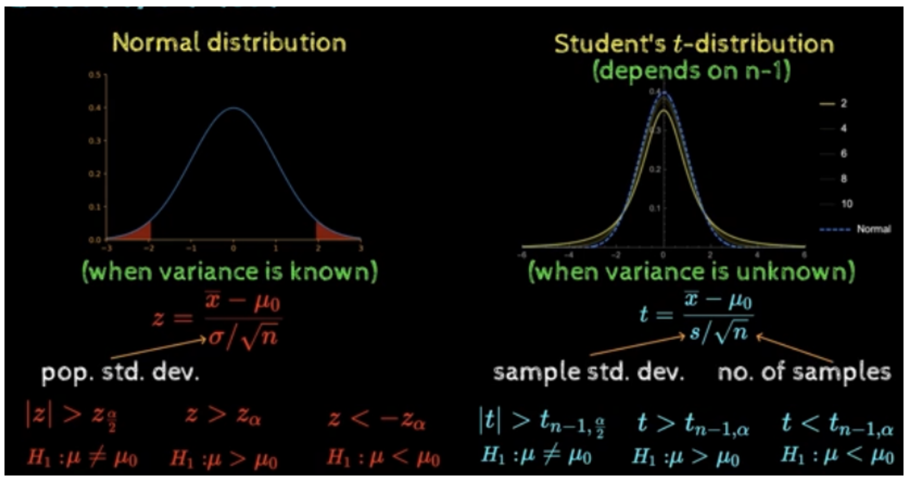
z-test vs t-test

- The population stddev. is not known so sample deviation is calculated and used.

- t test statistic is computed instead of z, where $t = \frac{\bar{x}-\mu_0}{s/\sqrt{n}}$. The variable has a t-distribution (instead of a normal distribution) and the critical regions are defined in the t-distribution.
- t-distribution is a family of distributions, thus, each sample size has a different distribution of freedom (sample size-1). a t table is used instead of a normal distribution table.



3 changes made when variance is unknown.



N(0,1) vs t-distribution.

Mean (unknown variance) - 2 case studies

Case study 1 : You have developed a new dialog stem and done a user study. You claim that the average rating given by the users is greater than 4 on a scale of 1 to 5. How do you prove this to your critiques? (use deviation = 0.5)

$H_1 : \mu > 4$ (bold claim) $H_0 : \mu \leq 4$ (status quo)

Sample : [4,3,5,4,5,3,5,5,4,2,4,5,5,4,4,5,4,5,4,5]

Mean $\bar{x}$ = 4.25 Standard deviation s = 0.85

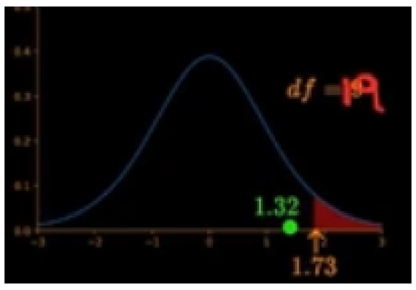
Test statistic $t = \frac{\bar{x}-\mu_0}{s/\sqrt{n}} = \frac{4.25-4}{0.85\sqrt{20}} = 1.32$

α = 0.05 (a 1-tailed test with 5% significance is done using t distribution with 9 degrees of freedom)

Decision rule : reject H_0 if $t > t_{n-1,\alpha}$
but 1.32 < 1.73

The null hypothesis cannot be rejected.

p value = $P(T > t) = 0.102$, there is a 10% chance of seeing such a sample



Case study 2. A company selling chips claims that each packet has 200gm of chips. You are skeptical of the claim and believe that on average each packet does not contain 200gm of chips. Prove your claim if the variance is not known

$H_1 : \mu \neq$ 200 (bold claim) $H_0 : \mu =$ 200 (status quo)

Sample : [193,212,174,200,195,195,194,198,181,203]

Mean $\bar{x}$ = 194.5 Standard deviation s = 10.7

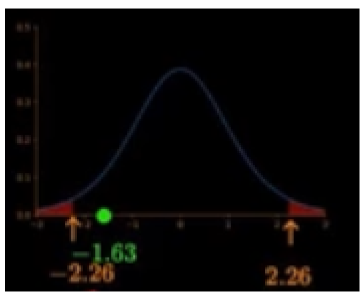
Test statistic $t = \frac{\bar{x}-\mu_0}{s/\sqrt{n}} = \frac{194.5-200}{107\sqrt{10}} = -1.63$

α = 0.05 (2-tailed test is done using a t-distribution with 9 degrees of freedom)

Decision rule : reject H_0 if $|t| > t_{n-1,\alpha/2}$
and | -1.63| < 2.26

The null hypothesis cannot be rejected.

p value = $P(|T| > t) = 0.138$



Proportion (p)

- Calculate the statistic $\bar{p}$ instead of mean $\bar{x}$. The s test statistic is calculated using $z = \frac{\bar{p}-p_0}{\sqrt{p(1-p)}/\sqrt{n}}$ and normal distribution is used.
- There is no case of unknown variance.

Case study 1 : You are teaching an online course and based on your internal surveys claim that 80% of the students like the course. A course review committee is skeptical about the claim and wants to test it.

$H_1 : p < 0.8$ $\quad\quad$ $H_0 : p \geq 0.8$

Sample : Interviewed 100 students and 82 said yes.

$\bar{p}$ = 0.82

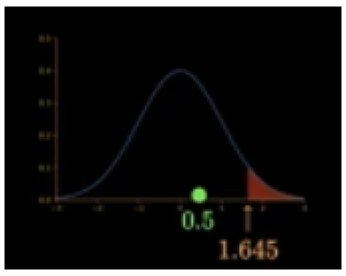
Test statistic $z = \frac{\bar{p}-p_0}{\sqrt{p(1-p)}/\sqrt{n}} = \frac{0.82-0.8}{\sqrt{0.8(1-0.8)}/\sqrt{100}} = 0.5$

$\alpha = 0.05$

Decision rule : reject H_0 is $z > z_\alpha$
but 0.5 < 1.645

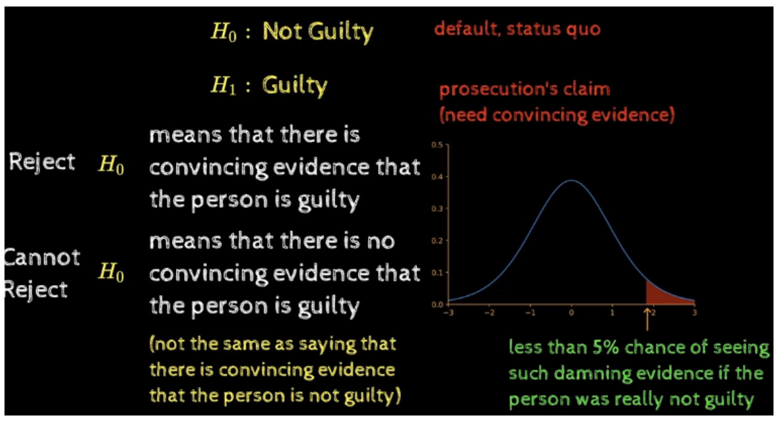
The null hypothesis cannot be rejected.

p value = $p(Z > z) = 0.308$



This implies that there is a 30% probability of seeing such a sample.

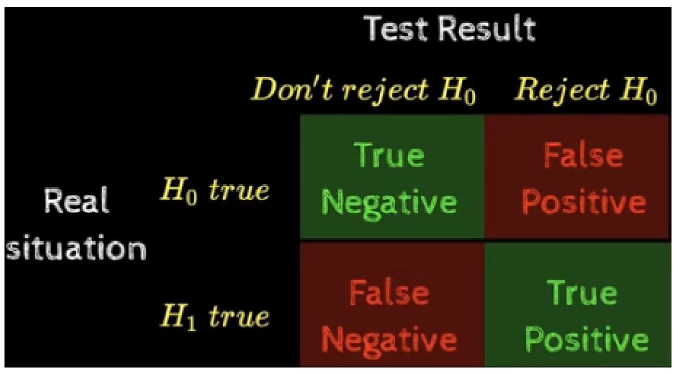
Type 1 & Type 2 Errors

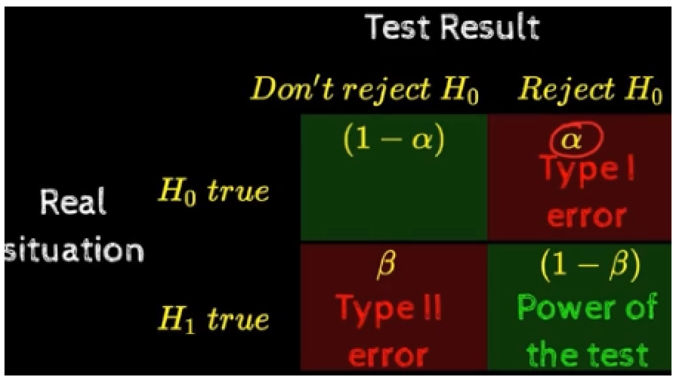


Courtroom analogy.

- There are four possible outcomes of the test (based on the reality and the test result):

1. True Negative : H_0 is true and not rejected. (not guilty, not rejected)
2. True positive :H_1 is true and H_0 is rejected. (guilty, rejected)
3. False positive : H_0 is true but rejected. **Type I error** (not guilty, rejected)
4. False negative: H_1 is true but H_0 is not rejected. **Type II error** (guilty, not rejected)





		Test Result	
		$Don't\ reject\ H_0$	$Reject\ H_0$
Real situation	$H_0\ true$	$(1-\alpha)$	α Type I error
	$H_1\ true$	β Type II error	$(1-\beta)$ Power of the test

- **Type I error** occurs when a person not guilty is found with extreme evidence and convicted. The probability of such an occurrence is α.
- **Type II error** occurs when a person is guilty but the null hypothesis is not rejected due to lack of evidence. Suppose the population mean μ_1 is different from what is inferred using null hypothesis μ_0, then μ_1 will have a distribution that has coincidence with the critical region of μ_0 and beyond. Thus, there is a possibility that the evidence lies beyond this critical region of μ_0 (not guilty curve) and in the critical region of μ_1(guilty curve). The probability of such occurrence is β.

Probability of type I and type II errors.

- $1-\beta$ is called the power of the test, where the guilty person is found and punished as the alternate hypothesis is proven. Ideally β should be small.

Two tailed and one tailed z-test

The variance is known

```
'''Example of a two tailed z-test'''
#given
mu_0 = 200
sigma = 10

#Collect Sample
sample = chip_weights
#sample = ball_bearing_radius

n = len(sample)
#Compute mean
mean = np.mean(sample)
#Compute test statistic
z = (mean - mu_0)/(sigma/sqrt(n))
#Set alpha
alpha = 0.05

z_critical = norm.ppf(1 - alpha/2)
p_value = 2 * (1.0 - norm.cdf(np.abs(z)))

print(z, z_critical)

p_value = float("{:.4f}".format(p_value))

if (np.abs(z) >  norm.ppf(1 - alpha/2)) :
  print ('Reject Null Hypothesis: p-value = ', p_value, ' alpha = ', alpha)
else :
  print ('Not enough evidence to reject Null Hypothesis: p_value = ', p_value, ' alpha = ', alpha)

>>> -1.7392527130926088 1.959963984540054
Not enough evidence to reject Null Hypothesis: p_value =  0.082  alpha =  0.05
```

```
'''Example of a one tailed z-test'''
#given
mu_0 = 4
sigma = 0.5

#Collect Sample
sample = user_ratings

n = len(sample)
#Compute mean
mean = np.mean(sample)
#Compute test statistic
z = (mean - mu_0)/(sigma/sqrt(n))
#Set alpha
alpha = 0.05

z_critical = norm.ppf(1 - alpha)
p_value = (1.0 - norm.cdf(np.abs(z)))

print(mean, z, z_critical)

p_value = float("{:.4f}".format(p_value))

if (np.abs(z) > z_critical) :
  print ('Reject Null Hypothesis: p-value = ', p_value, ' alpha = ', alpha)
else :
  print ('Not enough evidence to reject Null Hypothesis: p_value = ', p_value, ' alpha = ', alpha)

>>> 4.25 2.23606797749979 1.6448536269514722
Reject Null Hypothesis: p-value =  0.0127  alpha =  0.05
```

Two Tailed and One Tailed t-test

Variance is unknown.

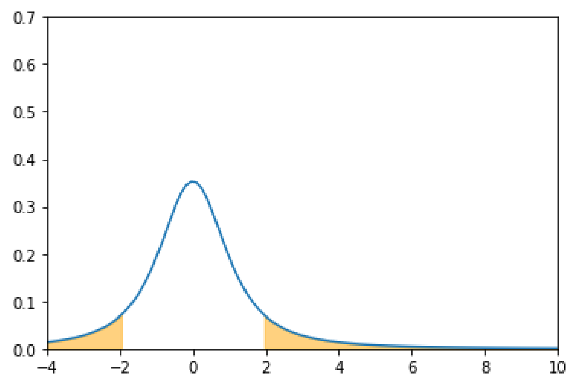
```
'''Example of a two tailed T-test'''
#Given
mu_0 = 200

#Collect Sample
sample = chip_weights

n = len(sample)
#Compute mean
mean = np.mean(sample)
stddev = np.std(sample, ddof=1)
#Compute test statistic
t_statistic = (mean - mu_0)/(stddev/sqrt(n))

#Set alpha
alpha = 0.05
t_critical = t.ppf(1 - alpha/2, n-1)
p_value = 2 * (1.0 - t.cdf(np.abs(t_statistic), n-1))

print(mean, stddev, t_statistic, t_critical)
p_value = float("{:.4f}".format(p_value))

if (np.abs(t_statistic) >  t_critical) :
  print ('Reject Null Hypothesis: p-value = ', p_value, ' alpha = ', alpha)
else :
  print ('Not enough evidence to reject Null Hypothesis: p_value = ', p_value, ' alpha = ', alpha)

>>> 194.5 10.67967956240053 -1.628562638916544 2.2621571627409915
Not enough evidence to reject Null Hypothesis: p_value =  0.1378  alpha =  0.05
```

```
'''Example of a one tailed T-test'''
#Given
mu_0 = 4

#Collect Sample
sample = user_ratings

n = len(sample)
#Compute mean
mean = np.mean(sample)
stddev = np.std(sample, ddof=1)
#Compute test statistic
t_statistic = (mean - mu_0)/(stddev/sqrt(n))
```

```python
#Set alpha
alpha = 0.05
t_critical = t.ppf(1 - alpha, n-1)
p_value = (1.0 - t.cdf(np.abs(t_statistic), n-1))

print(n, t_statistic, t_critical, stddev)
p_value = float("{:.4f}".format(p_value))

if (np.abs(t_statistic) >  t_critical) :
  print ('Reject Null Hypothesis: p-value = ', p_value, ' alpha = ', alpha)
else :
  print ('Not enough evidence to reject Null Hypothesis: p_value = ', p_value, ' alpha = ', alpha)

>>> 20 1.3142574813455419 1.729132811521367 0.8506963092234007
Not enough evidence to reject Null Hypothesis: p_value =  0.1022  alpha =  0.05
```

Plotting distributions

```python
x_min =  -4
x_max = 10

mean = 0
std = 1

x = np.linspace(x_min, x_max, 100)

#y = norm.pdf(x, mean, std)
y = t.pdf(x, 2)
#y3 = chi2.pdf(x, 4)

ax = plt.gca()

plt.xlim(x_min,x_max)
plt.ylim(0,0.7)

plt.plot(x,y)

x1 = np.linspace(-4, -1.96, 100)
y1 = t.pdf(x1, 2)
plt.fill_between(x1, y1, color='orange', alpha=0.5)

x2 = np.linspace(1.96, 10, 100)
y2 = t.pdf(x2, 2)
plt.fill_between(x2, y2, color='orange', alpha=0.5)

plt.savefig("normal_dist.png", dpi=400, transparent=True)
plt.show()
```



Chi-Square test of independence

- Degrees of freedom = (rows - 1)(cols - 1)

Case study 1 : Is there a relationship between gender and preference of OS?

H_1 : they are dependent

H_0 : they are independent

Sample is collected and sample (observed) frequency is calculated.

	Male	Female	Total
iOS	25	40	65
Windows	60	70	130
Linux	20	30	50
Total	105	140	245

Observed frequencies.

If the null hypothesis was true, the the expected frequencies would have had same fraction of male and female for each OS:

	Male	Female	Total
iOS	27.85	37.15	65
Windows	55.70	74.30	130
Linux	21.50	28.50	50
Total	105	140	245

$\frac{65}{245} * 105$

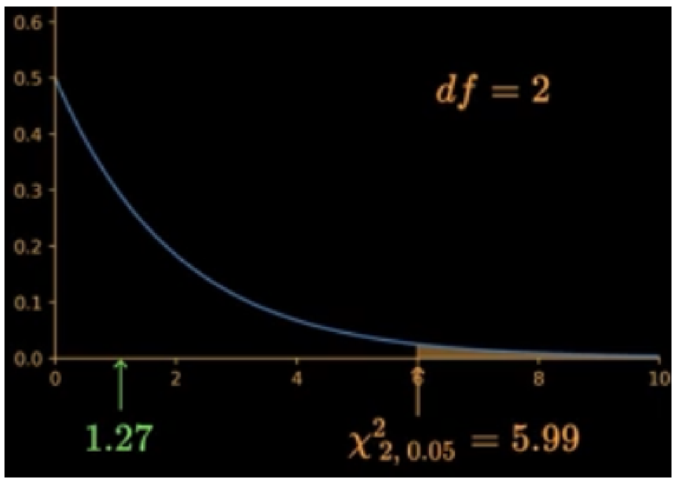
$\frac{65}{245} * 140$

Expected frequencies.

Compute χ^2 test statistic : $\chi^2 = \sum \frac{(o-e)^2}{e} = 1.27$

Using χ^2 distribution with df=2 and $\alpha = 0.05$

Decision rule : reject H_0 if $\chi^2 > \chi^2_{df,\alpha}$
but 1.27<5.99



The null hypothesis cannot be rejected based on the sample taken.

p value = $P(\chi^2_2 > 1.27) = 0.52$

Case study 2 : Is there a relationship between marital status (married, single, divorced) of patients being treated for depression and the severity of their condition (severe, normal, mild)

H_1 : they are independent

H_0 : they are dependent

Sample is collected and sample (observed) frequency is calculated.

	Married	Single	Divorced	Total
Severe	22	16	19	57
Normal	33	29	14	76
Mild	14	9	3	26
Total	69	54	36	159

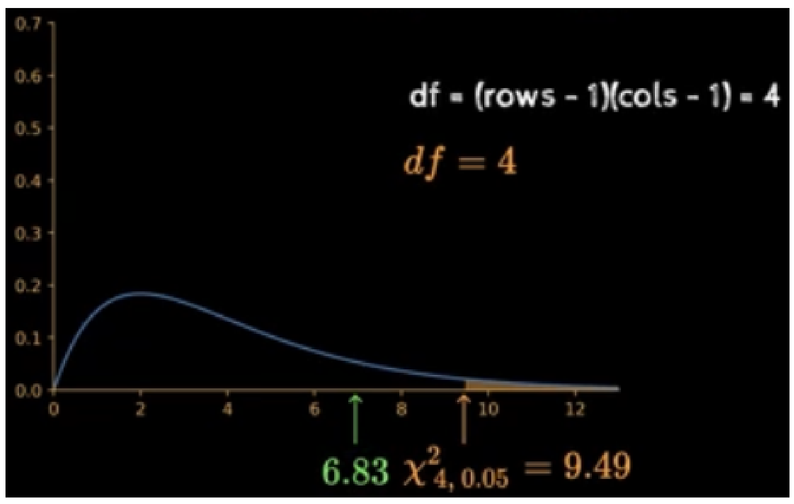
Calculating expected frequencies:

	Married	Single	Divorced	Total
Severe	24.75	19.35	12.9	57
Normal	33.0	25.8	17.2	76
Mild	11.3	8.8	5.9	26
Total	69	54	36	159

Compute χ^2 test statistic : $\chi^2 = \sum \frac{(o-e)^2}{e} = 6.83$

Using χ^2 distribution with df=4 and $\alpha = 0.05$

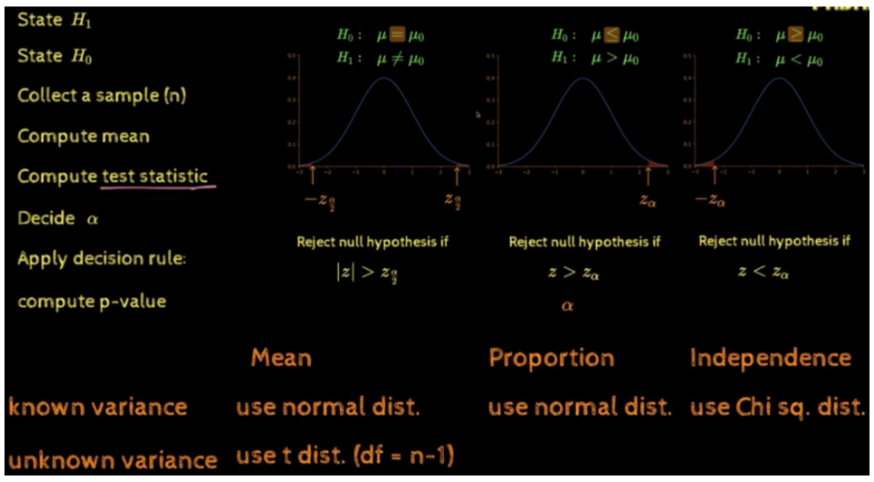
Decision rule: reject H_0 if $\chi^2 > \chi^2_{df,\alpha}$
but 6.83<9.49



The null hypothesis cannot be rejected based on the sample taken.

p value = $P(\chi_2^2 > 1.27) = 0.145$

Summary Slide



MCQ

1. Rejection of the null hypothesis is a conclusive proof that the alternative hypothesis is
    1. True
    2. False
    - ▼ **Neither**

      The rejection of the null hypothesis in favor of the alternative hypothesis cannot be taken as conclusive proof that the alternative hypothesis is true, but rather as a piece of evidence that increases one's belief in the truth of the alternative hypothesis.

2. There are two types of errors in hypothesis testing Type I and Type II. Type II error is committed when
    1. Null hypothesis is rejected and the alternate hypothesis is true.
    2. Null hypothesis is rejected when it is true
    - ▼ **Null hypothesis is accepted when it is not true**

      Type I error means that the null hypothesis that was true has been rejected. The type II error is opposite, i.e. the false null hypothesis that should have been rejected has in fact been accepted in favour of the alternative hypothesis that is true.

3. $\alpha = 0.05$ implies
    1. 5% confidence that the results have not occurred by chance
    2. 95% confidence that the results have occurred by chance.
    - ▼ 95% confidence that the results have not occurred by chance.

      If an analyst states that the results are significant at the 5% level then what they are saying is that there is a 5% probability that the sample data values collected have occurred by chance. An alternative view is to use the concept of a confidence interval. In this case we can observe that we are 95% confident that the results have not occurred by chance

4. one or two tailed tests determine
    1. if the extreme values of the samples need to be rejected
    2. The number of possible conclusions the hypothesis has
    - ▼ If the region of rejection is located in one or two tails of the distribution.

      The statement that defines the problem will state if the tested value should be inside a range, or greater/smaller than some specified value. From this statement we'll know if one or two tail test is needed. The test will determine if the region of rejection is located in one tail or two tails of the sampling distribution.

5. The level of significance can be viewed as the amount of risk that an analyst will accept when making a decision
    1. False

- **True**

  If level of significance is, for example 0.05, then this means that 5% of the cases might be randomly generated and that an analysis is certain that 95% of the cases comply with the hypothesis. In this respect, the level of significance can be viewed as the risk factor.

6. p-value in hypothesis testing represents
    1. probability of not rejecting the null hypothesis, given the observed results
    2. probability that null hypothesis is true, given the observed results
    3. probability that the observed results are statistically significant, given that the null hypothesis is true
    4. **probability of observing results as extreme or ore extreme than currently observed, given that the null hypothesis is true.**
7. Null hypothesis and alternative hypothesis are statements about
    1. **population parameters**
    2. sample parameters
    3. sample statisitics
    4. depends on the use-case
8. A result is called statistically significant when
    1. the null hypothesis is true
    2. the alternative hypothesis is true
    3. **the p-value is less or equal to the significance level.**
    4. the p-value is larger than the significance level.
9. A plausible way to determine if a statistically significant difference in two means is of practical importance is to
    1. **find a 95% confidence interval and notice the magnitude to the difference**
    2. repeat the study with the same sample size and see if the difference is statistically significant again
    3. see if the p-value is extremely small
    4. see if the p-value is extremely large